

## 인과관계 분석의 어려움

---

# Titanic Tragedy



- What happened to the Titanic in that night?

1912년 4월 10 일, 영국의 Southampton을 출발, 뉴욕으로 첫 항해를 하던 Titanic호가 4월 15 일 대서양에서 빙산과 충돌하고 침몰, 1500 여명의 사상자가 발생.

승객들을 구하거나 품위있게 죽은 영웅들도 있었다. 이들 중 일부는 영화 타이타닉에도 등장한다.

## 타이타닉호의 비극에서 알려진 미담

- 에드워드 존 스미스 선장은 구명보트에 탈 수 있음에도 불구하고 끝까지 승객들을 지도하다가 끝까지 배에 남아 죽었다. 참고로 스미스 선장은 낮은 계층 출신이었으나 그 기품과 카리스마 덕분에 화이트 스타 라인의 고객 중에도 오로지 스미스 선장의 배만을 타는 사람도 많았다고 한다. 그는 원래 1911년에 은퇴할 예정이었지만 회사측의 설득으로 마지막으로 한 번 더 출항하기로 한 것이 바로 타이타닉호의 첫 항해였다. 항해사 절반 또한 죽었다.
- 배의 설계자인 토머스 앤드루스는 승객들의 구명보트 탑선과 뜰 만한 물건들을 던지는 것을 돕다가 1등실 흡연실에 조용히 들어가서 배와 함께 최후를 맞이했다. 이때 흡연실에 고귀하게 남기로 한 사람은 앤드루스 뿐만이 아니라 다른 (1등실) 승객들도 있었다. 어떤 승객들은 카드 게임을 계속했으며 당대 저명한 언론인이었던 윌리엄 T. 스티드는 조용히 독서를 하고 있었다. 참고로 윌리엄 스티드는 과거에 헤이그 특사를 도와준 유일한 외국인 이었다. 이들 또한 모두 배와 함께 죽었다.

## 타이타닉호의 미담?



## 승무원들이 희생정신을 발휘하여 승객들을 적극 구조?

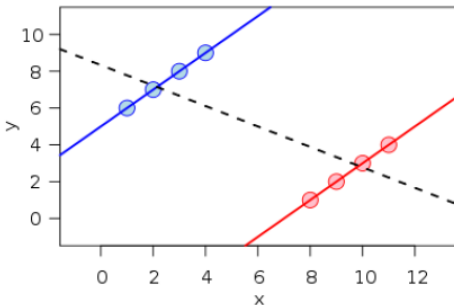
	생존자	사망자	생존율
1등실 성인 승객	197	122	61.7%
2등실 성인 승객	94	167	36.0%
3등실 성인 승객	151	476	24.1%
승무원	212	673	24.0%

- 승무원들의 생존율이 가장 낮음. (승무원들의 희생 정신?)
- 승무원들의 생존율은 3등석 승객과 비교하여 낮지 않음.
- Simson's paradox

	어린이				성인				생존율		여성 비율
	남자		여자		남자		여자		남자	여자	
	생존	사망	생존	사망	생존	사망	생존	사망	%	%	
1등실	5	0	1	0	57	118	140	4	32.6	97.2	45.0
2등실	11	0	13	0	14	154	80	13	8.3	86.0	35.0
3등실	13	35	14	17	75	387	76	89	16.2	46.1	26.3
승무원	0	0	0	0	192	670	20	3	22.3	86.9	2.6

# Simson's paradox

- 승무원들이 생존율이 낮게 나온 것은 승무원 중에서 남성에 비하여 생존율이 높은 여성의 비율이 낮기 때문
- 다른 변수의 영향을 무시할 경우 흔히 발생



## 숨겨진 변수의 영향

- Sex Bias in Graduate Admissions:

In 1973, the University of California-Berkeley was sued for sex discrimination.

The numbers looked pretty incriminating: the graduate schools had just accepted 44% of male applicants but only 35% of female applicants.

Applicants	Outcome				Difference	
	Observed		Expected			
	Admit	Deny	Admit	Deny	Admit	Deny
Men	3738	4704	3460.7	4981.3	277.3	-277.3
Women	1494	2827	1771.3	2549.7	-277.3	277.3

- When researchers looked at the evidence, though, they uncovered something surprising.

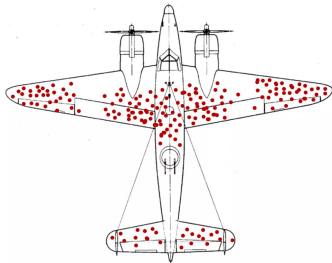
- In 1973, UC Berkeley had 101 different graduate departments, but for simplicity, we will consider only the six largest departments (which collectively account for 4486 of the applicants.)
- Based on this analysis, do you now believe that there was gender bias in graduate admissions at the University of California at Berkeley in 1973?

	Male		Female	
	Admitted	Denied	Admitted	Denied
A	512(62%)	313	89(83%)	19
B	313(60%)	207	17(68%)	8
C	120(37%)	205	202(34%)	391
D	138(33%)	279	131(35%)	244
E	53(28%)	138	94(24%)	299
F	22(6%)	351	24(7%)	317



## 데이터는 중요하다. 그러나 데이터가 모든 것을 말해주지는 않는다

- 2차대전 당시, 영국 주둔 미공군이 독일 본토 폭격 임무를 마치고 돌아온 폭격기를 조사한 결과 날개와 중앙 동체, 그리고 꼬리날개 부분의 피해가 크다는 사실을 발견
- 즉시 피해를 줄이기 위해 날개와 동체 부분에 방탄재를 추가하려는 시도가 진행



<http://flip.it/2lB6Yz>

- 그러나 한 항공기 정비사가 이에 반기를 들었다.
- “날개가 아닌 엔진과 조종석에 방판재를 강화해야 합니다. 엔진과 조종석이 피탄된 폭격기들은 귀환하기 못했기 때문입니다.”
- 여기에서 생존 편향(survivorship bias)의 오류라는 말이 파생.
  - 어떤 현상에 대한 결과를 놓고 그 원인과 대책을 수립하려 할 때, 결과에 대한 피상적인 집착이 오히려 그릇된 결과를 가져올 수 있다는 의미
- 온라인 비즈니스에서도 비슷한 현상이 목격된다.
  - 특히 최근 빅데이터 분석이나 데이터 마케팅이라는 이름으로 다양한 데이터를 통해 비즈니스 통찰을 얻으려는 시도가 시행되고 있다.
  - 그러나 여기에도 함정은 있다. 너무 많은 데이터와 혼란, 그리고 시간과 비용이 소모된다는 점이다.
  - 데이터 마케팅 자체가 고가의 마케팅 포장 도구일 뿐이라는 비판도 적지 않다.

- 그렇다면 반론을 제기한 항공기 정비사는 어떻게 생존 편향의 오류에서 벗어날 수 있었을까?
  - 보고서에 적힌 숫자가 아닌 격납고에 놓인 폭격기를 보면 한 눈에 무엇이 문제인지 알 수 있다.
  - 기체에 박힌 수많은 총탄 구멍에도 불구하고 엔진과 조종석에는 피탄 흔적이 없다. 이 때문에 조종사는 살아서 귀환할 수 있었다.
  - 정비사는 자신이 해야 할 일이 무엇인지 직관적으로 알 수 있었다. 바로 엔진과 조종석 주위에 방탄재를 더 보강하는 것이다.
- 데이터는 중요하다. 그러나 데이터가 모든 것을 말해주지는 않는다
- 때론 보고서나 대시보드 화면 밖 현장에 답이 있다.
- 데이터를 어떻게 분석할 것인가가 중요

